引用文献4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 June 2002 (27.06.2002)

PCT

(10) International Publication Number
WO 02/50772 A1

(51) International Patent Classification[7]: G06T 9/00, H04N 7/26

(21) International Application Number: PCT/AU01/01660

(22) International Filing Date:
21 December 2001 (21.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PR 2225 21 December 2000 (21.12.2000) AU

(71) Applicant *(for all designated States except US)*: UNISEARCH LIMITED [AU/AU]; Rupert Myers Building, Gate 14, Barker Street, Sydney, New South Wales 2052 (AU).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: TAUBMAN, David [AU/AU]; 20 Wingate Avenue, Eastwood, New South Wales 2122 (AU).

(74) Agent: GRIFFITH HACK; GPO Box 4164, Sydney, New South Wales 2001 (AU).

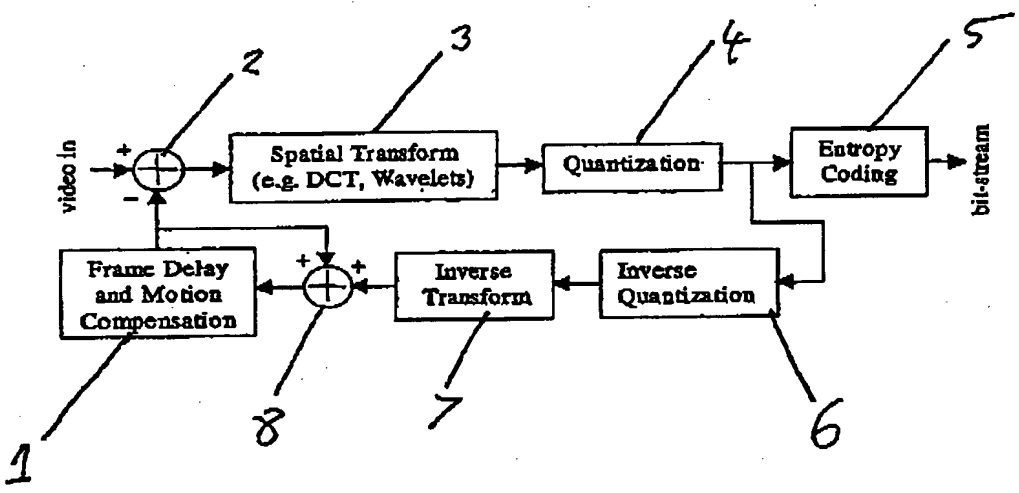
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR SCALABLE COMPRESSION OF VIDEO



(57) Abstract: The present invention relates to a method and apparatus for producing a fully scalable compressed representation of video sequences, so that they may be transmitted over networks, such as the Internet, for example. Because the signal is scalable, users receiving the signal can obtain the signal at the appropriate resolution and quality that their system will handle or that they desire. The invention implements a "motion compensated temporal wavelet transform" in order to enable compression of the video in a scalable fashion while still taking advantage of inter-frame redundancy in a manner which is sensitive to scene and camera motion. The motion compensated temporal wavelet transformed is implemented by decomposing the video sequence into a set of temporal frequency bands and then applying a sequence of motion compensated lifting operations to alternatively update an odd frame sub-sequence based upon an even sub-sequence and vice versa in a manner which is sensitive to motion.

WO 02/50772 A1

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND APPARATUS FOR SCALABLE COMPRESSION OF VIDEO

Field of the Invention

The present invention relates generally to the processing of video signals so that video may be transmitted in a compressed form and, particularly, but not exclusively, relates to processing of video signals so that the video can be efficiently transmitted in a scalable form.

Background of the Invention

Currently, most video content which is available over computer networks such as the Internet must be pre-loaded in a process which can take many minutes over typical modem connections, after which the video quality and duration can still be quite disappointing. In some contexts video streaming is possible, where the video is decompressed and rendered in real-time as it is being received; however, this is limited to compressed bit-rates which are lower than the capacity of the relevant network connections. The most obvious way of addressing these problems would be to compress and store the video content at a variety of different bit-rates, so that individual clients could choose to browse the material at the bit-rate and attendant quality most appropriate to their needs and patience. Approaches of this type, however, do not represent effective solutions to the video browsing problem. To see this, suppose that the video is compressed at bit-rates of R, 2R, 3R, 4R and 5R. Then storage must be found on the video server for all these separate compressed bit-streams, which is clearly wasteful. More importantly, if the quality associated with a low bit-rate version of the video is found to be insufficient, a complete new version must be downloaded at a higher bit-rate; this new bit-stream must take longer to

download, which generally rules out any possibility of video streaming.

To enable real solutions to the remote video browsing problem, scalable compression techniques are required. Scalable compression refers to the generation of a bit-stream which contains embedded subsets, each of which represents an efficient compression of the original video with successively higher quality. Returning to the simple example above, a scalable compressed video bit-stream might contain embedded sub-sets with the bit-rates of R, 2R, 3R, 4R and 5R, with comparable quality to non-scalable bit-streams with the same bit-rates. Because these subsets are all embedded within one another, however, the storage required on the video server is identical to that of the highest available bit-rate. More importantly, if the quality associated with a low bit-rate version of the video is found to be insufficient, only the incremental contribution required to achieve the next higher level of quality must be retrieved from the server. In a particular application, a version at rate R might be streamed directly to the client in real-time; if the quality is insufficient, the next rate-R increment could be streamed to the client and added to the previous, cached bit-stream to recover a higher quality rendition in real time. This process could continue indefinitely without sacrificing the ability to display the incrementally improving video content in real time as it is being received from the server.

A major problem, however, is that highly efficient scalable video compression algorithms have not existed, either in practice or in the academic literature. Efficient scalable image compression algorithms have existed for some time, of which the most well known examples are the so-called embedded zero-tree algorithms initially proposed by (J. Shapiro, "An embedded hierarchical image coder using zerotrees of wavelet coefficients", *Data Compression Conference (Snowbird,*

Utah), PP. 214-223, 1993 and later enhanced by A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees" *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, PP. 243-250, June 1996. In fact, many of the algorithms advanced for scalable video compression are essentially scalable image compression schemes, applied independently to the successive frames of the video sequence, see S. McCanne, M. Vetterli and V. Jacobson, "Low-complexity video coding for receiver-driven *Layered Multicast*," *IEEE Journal on Selected Areas in Communications*, vol. 15, August 97, PP. 983-1001. In order to compete with the efficiency of non-scalable techniques, however, it is essential that inter-frame redundancy be exploited in a manner which is sensitive to scene and camera motion.

Motion Compensated Prediction (MCP) is by far the most popular approach to exploit inter-frame redundancy for video compression. Figure 1 illustrates the salient features of MCP compression, upon which key standards such as MPEG-1, MPEG-2, MPEG-4 and H.263, all rely. Rather than compressing each frame of the video sequence separately, the spatial transform, quantisation and entropy coding elements common to image compression algorithms are applied to the difference between the current frame and a prediction of the frame formed by applying a motion compensation algorithm to the pixels in the previous frame, as reconstructed by the decompressor.

Figure 1 is a schematic block diagram of a prior art arrangement for compressing video using a motion compensated prediction (MCP) feedback loop. It will be appreciated that the blocks shown in the diagram are implemented by appropriate hardware and/or software.

Block 1 is a frame delay and motion compensator. This stores the decoded version of a previous frame, using motion information (usually explicitly transmitted with the compressed data stream) to form a prediction of the current frame. The subtracter 2 subtracts the motion

compensated predictor, produced by block 1, from the current video frame. The spatial transform block 3 decomposes the prediction residual frame produced by block 2 into separate components for coding. The separate components usually correspond to different spatial frequencies and are less correlated with each other than are the original samples of the prediction residual frame. The quantisation block 4 approximates each of the transform coefficients, by a number of representative values, identified by labels (usually integers) which are readily coded. This step is a precursor to coding.

Block 5 is an entropy coder which produces a bit-stream which efficiently represents the quantisation labels produced by block 4, and which can be transmitted over a network.

The inverse quantisation block 6 uses the labels produced by block 4 to reconstruct representative values for each of the transform coefficients which were quantised by block 4.

The inverse transform block 7 is the inverse of the spatial transform operator.

Block 8 adds the decoded prediction residual recovered from block 7 to the predictor itself, thereby recovering a copy of the decoded video frame, identical to that which should be available at a decompressor.

MCP relies upon predictive feedback. It requires knowledge of the pixel values which would be reconstructed by the decompressor for previous frames. In a scalable setting, this knowledge is unavailable, because the pixel values reconstructed by the decompressor depend upon the particular subset of the embedded bit-stream which is actually received and decompressed. This problem has been primarily responsible for hampering the development of efficient highly scalable video compressors. MPEG-2 allows some small degree of scalability, but the useful extent of this capability is limited to two or three different quality subsets, and even then with significant

loss in efficiency. Although MPEG-4 claims to be highly scalable, this claim is to be interpreted in regard to so-called "object scalability" which is limited to the number of "objects" in the scene, dependent upon appropriate segmentation algorithms and unable to provide smooth increments in video quality as a function of the bandwidth available to a remote client. Otherwise, MPEG-4 is constrained by the fundamental inappropriateness of MCP, upon which it relies to exploit inter-frame redundancy.

Two notable scalable video compression algorithms which have been proposed are those of J. Ohm, "Three dimensional sub-band coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, pp. 559-571, September 1994 and D. Taubman and A. Zakhor "Multi-rate 3-D sub-band coding of video," *IEEE Trans. Image Processing*, vol. 3, pp. 572-588, September 1994. In both cases, the idea is to use three-dimensional separable sub-band transforms without any predictive feedback, after first temporally shifting the video frames, or parts thereof, so as to improve the alignment of spatial features prior to application of the 3-D transform. Although these schemes work well for simple global translation, their performance suffers substantially when scene motion is more complex.

Summary of the Invention

In accordance with a first aspect, the present invention provides a method of compressing a video sequence to produce a compressed video signal, including the steps of forming the video sequence into an input signal, decomposing the input signal into a set of temporal frequency bands, in a manner which exploits motion redundancy and, applying scalable coding methods to the decomposed input signal to generate a scalable compressed video signal.

One of the important features of the present invention is that motion redundancy is exploited directly

by the decomposition into temporal frequency bands, thereby avoiding the need for predictive feedback. The decomposed input signal is therefore able to be coded in a scalable fashion.

Preferably, the step of decomposing the input signal comprises the further step of decomposing the temporal frequency bands into spatial frequency bands, to provide spatio-temporal frequency bands. As discussed later, it is sometimes desirable to rearrange the steps of temporal and spatial decomposition in a manner which allows efficient access to the video at various reduced sizes (spatial resolutions).

Motion compensation is preferably incorporated by separating the input signal into even and odd indexed frame sub-sequences, and applying a sequence of "motion compensationed lifting" operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa. The low temporal frequency band is preferably further decomposed for a predetermined number of decomposition levels by recursively applying the lifting operations.

After further decomposition of the temporal bands into spatial frequency bands, the resulting spatio-temporal frequency bands are subjected to scalable coding methods to generate the scalable compressed video signal. Preferably, the samples belonging to each spatio-temporal frequency band are partitioned into blocks, such that each block is coded independently of the other, with the spatial and temporal dimensions of said blocks selected in such a manner as to provide efficient access into regions of interest within the video.

Preferably, the blocks are coded in an embedded manner, providing incremental contributions to embedded subsets in a digital bit stream. In this way, the embedded subsets can be selectively decoded by a computing system to enable a user to view the desired part of the video, at the desired available (scaled) resolution and

quality.

In the preferred embodiment, the scalable bit-stream therefore contains distinct subsets corresponding to different intervals in time, different resolutions (both temporal and spatial) and different levels of quality; a client can interactively choose to refine the quality associated with specific time segments which are of the greatest interest. Further, the scalable bit-stream also may contain distinct subsets corresponding to different spatial regions and clients can then interactively choose to refine the quality associated with specific spatial regions over specific periods of time, according to their level of interest. In a training video, for example, a remote client in a network could interactively "revisit" certain segments of the video and continue to stream higher quality information for these segments from a network server, without incurring any delay.

The provision of a highly scalable compressed video signal, in accordance with the present invention, can facilitate interactive browsing of video signals by a computer on a network, such as the Internet, for example.

In accordance with a second aspect of the present invention, there is provided a method of providing for interactive remote browsing of compressed digital video signals, comprising the steps of:

making available to a network, by way of a server computing system connected to the network, a scalable compressed signal representing a video sequence,

selecting, by way of a client computing system connected to the network, a portion of the scalable compressed signal for decompression and viewing as a video sequence, monitoring the choice of selection by the client, and selecting, by way of the client, a further portion of the scalable compressed signal of interest to the client.

Preferably, the client includes a cache which can store the portion of the signal which the client has

viewed so far. If the client wishes to view a further portion of the signal associated with the same area of video (which may include temporal area as well as spatial area), all they need is a sufficient further portion of the signal to, for example, increase the resolution or quality and provide a more accurate rendition of the video. They do not need to be resent the entire signal. Preferably, the server also includes a cache monitoring means, which monitors the status of the client cache, so that the server knows what portion of the signal the client has already received and can determine what further portion of the scaled signal is required.

In accordance with a third aspect of the present invention, there is provided a method in accordance with the second aspect of the present invention wherein the scalable compressed signal is compressed by the method of the first aspect of the present invention.

In accordance with a fourth aspect, the present invention provides a method of compressing a video sequence to produce a compressed video signal, including the steps of decomposing the video sequence into a set of temporal frequency side bands by separating the video sequence into even and odd indexed frame sub-sequences, and applying a sequence of motion compensated lifting operations to alternately update the odd frame-sequence based on the even frame-sequence and vice versa in a manner which is sensitive to motion, and applying coding methods to generate a compressed video signal.

In accordance with a fifth aspect, the present invention provides a system for compressing a video sequence to produce a compressed video signal, the system comprising means for forming the video sequence into an input signal, means for decomposing the input signal into a set of temporal frequency bands, in a manner which exploits motion redundancy, and a means for applying a scalable coding method to the decomposed input signal to generate a scalable compressed video signal.

In accordance with a sixth aspect, the present invention provides a system for providing for interactive remote browsing of compressed digital video signals comprising:

a server computing system arranged to be connected to a network and including means for providing a scalable compressed signal representing a video sequence for distribution on the network to a client computing system,

a client computing system, including selection means for selecting a portion of the scalable compressed signal, a decompressor for decompressing the scalable signal for viewing as a video sequence, the selection means being arranged to select further portions of the scalable signal relating to regions of interest of the client.

In accordance with a seventh aspect, the present invention provides an apparatus for compressing a video sequence to produce a compressed video signal, comprising means for decomposing the video sequence into a set of temporal bands, and separating the input signal into even and odd index frame sub-sequences and applying a sequence of motion compensated lifting operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa in a manner which is sensitive to motion.

In accordance with an eighth aspect, the present invention provides a method in accordance with the first aspect of the invention, comprising the steps of applying method steps which are the inverse of the steps of a method in accordance with the first aspect of the invention.

In accordance with a ninth aspect, the present invention provides a decompressor system including means for applying a method in accordance with the eighth aspect of the invention.

In accordance with a tenth aspect, the present invention provides a computer program arranged, when loaded onto a computing system, to control the computing

system to implement a method in accordance with the first aspect of the invention.

In accordance with an eleventh aspect, the present invention provides a computer readable medium, arranged to provide a computer program in accordance with the tenth aspect of the invention.

In accordance with a twelfth aspect, the present invention provides a system where the client includes a cache arranged to store portions of the scalable signal and the server includes a cache monitor arranged to monitor the client to enable the server to determine what portions of the signal are stored in the client cache, whereby the server, on request of the client for a region of interest video, may send the client only the further portions of signal they require to reproduce the region of interest.

Brief Description of the Drawings

Features and advantages of the present invention will become apparent from the following description of embodiments thereof, by way of example only, with reference to the accompanying drawings, in which:

Figure 1 is a schematic block diagram illustrating compression and coding of a video signal utilising Motion Compensated Prediction feedback loop (prior art);

Figure 2 is a schematic block diagram of a video compression system in accordance with an embodiment of the present invention;

Figure 3 is a schematic block diagram illustrating a system for interactive remote browsing of scalable compressed video, in accordance with an embodiment of the present invention, and

Figure 4 is a schematic block diagram of a further embodiment of a compression system in accordance with the present invention, implementing spatial resolution scalable compression.

Detailed Description of Preferred Embodiments of the Invention

As mentioned already, the most fundamental obstacle to the development of effective scalable video compression technology is the difficulty of exploiting inter-frame redundancy within a highly scalable framework. In place of MCP, the most successful approaches in the past have focused on the use of the Haar Wavelet in the temporal domain, generating temporal low- and high-pass sub-bands corresponding to the average and difference of adjacent frames. The Wavelet transform is iterated along the low-pass channel, generating further low- and high-pass sub-bands from the low-pass channel in a recursive manner. To minimise memory requirements, this is usually continued for only two or three iterations so that the lowest frequency temporal sub-band might represent the average of 8 adjacent frames.

In order to compensate for motion, the individual frames can be pre-warped in a manner which aligns common spatial features in adjacent frames. This is the key idea in D. Taubman and A. Zakhor, "Multi-rate 3-D sub-band coding of video," *IEEE Trans Image Processing*, vol. 3, pp. 572-588, September 1994 and is related to the approach of J. Ohm, "Three dimensional sub-band coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, pp. 559-571, September 1994; however, the warping must be invertible or very nearly so, in order to ensure efficient compression and ensure that the original video sequence can be recovered in the absence of quantisation error introduced during subsequent coding stages. The invertibility requirement limits the spatial warpings and hence the types of scene motion which can be captured, to translations and skew approximations to rotation. It is impossible to capture the local expansions and contractions associated with camera zoom and scene object

motion through warping individual frames, without violating the invertibility requirement. In some proposed schemes e.g. J. Tham, S. Ranganath and A. Kassim, "Highly scalable wavelet-based video code for very low bit-rate environment," *IEEE Journal on Selected Areas in Communications*, vol. 16, January 1998, pp. 12-27, the invertibility requirement is deliberately violated, so that high quality reconstruction is impossible.

In order to effectively exploit inter-frame redundancy under complex scene motion, the temporal Wavelet transform and motion compensation operations are performed jointly in the present invention. We call the resulting transform a Motion Compensated Temporal Wavelet Transform (MCTWT). The implementation of the MCTWT is based upon the lifting mechanism for realising Wavelet transforms, see R. Calderbank, I. Daubechies, W. Sweldens and B. Yeo, "Wavelet transforms that map integers to integers," *Applied and Computational Harmonic Analysis*, vol. 5, pp. 332-369, July 1998.

In the specific case of the Haar Wavelet, the invention is most easily understood as a generalisation of the well-known S-Transform. Specifically, let $x_k[m,n]$ denote the sequence of frames from the video sequence. The lifting steps in the MCTWT transform the even and odd frame sub-sequences, $x_{2k}[m,n]$ and $x_{2k+1}[m,n]$, into low and high-pass temporal sub-bands, respectively. Let $W_{i,j}$ denote the motion compensated mapping of frame $x_i[m,n]$, onto the coordinate system of frame $x_j[m,n]$, so that $(W_{i,j}(x_i))[m,n] = x_j[m,n], \forall m,n$. If $i<j$, the mapping $W_{i,j}$ corresponds to what is commonly known as forward motion compensation, as used in elementary motion compensated prediction algorithms. If $i>j$, the mapping $W_{i,j}$

corresponds to what is commonly known as backward motion compensation, as used in bi-directional motion compensated prediction algorithms. In various embodiments of the invention these motion compensated mappings may correspond to conventional block based motion compensation, whose underlying motion model is piecewise constant, or more complex mesh based motion compensation, whose underlying motion model is a continuous interpolation of motion vectors on a control grid. Other motion compensation mappings may also be applied in various embodiments of the invention, and the invention is not limited to any particular type. Within the context of the MCTWT, the high-pass sub-band's frames are given by

$$h_k[m,n] = x_{2k+1}[m,n] - (W_{2k,2k+1}(x_{2k}))[m,n] \quad (1)$$

while the low-pass temporal sub-band samples are given by

$$l_k[m,n] = x_{2k}[m,n] + \frac{1}{2}(W_{2k+1,2k}(h_k))[m,n] \quad (2)$$

Evidently, when there is no motion, so that $W_{i,j}$ is the identity operator for all i,j, this sequence of two lifting steps reduces to the Haar Wavelet transform (up to a scale factor).

The approach is readily extended to more complex Wavelet transforms, involving more lifting steps, or lifting steps with larger kernels. A further example is based upon the well-known bi-orthogonal 5-3 Wavelet transform, both for completeness and because this example can be preferable to the Haar example given above for some embodiments of the invention. In this case, the high-pass MCTWT samples become

$$(3)\quad h_k[m,n] = x_{2k+1}[m,n] - \frac{1}{2}\left\{\left(W_{2k,2k+1}(x_{2k})\right)[m,n] + \left(W_{2k+2,2k+1}(x_{2k+2})\right)[m,n]\right\}$$

while the low-pass temporal sub-band samples are given by

$$(4)\quad l_k[m,n] = x_{2k}[m,n] + \frac{1}{4}\left\{\left(W_{2k-1,2k}(h_{k-1})\right)[m,n] + \left(W_{2k+1,2k}(h_k)\right)[m,n]\right\}$$

Again, when there is no motion, this sequence of two lifting steps reduces to the 5-3 bi-orthogonal Wavelet transform. It is interesting to note that the high-pass sub-bands in the case of the Haar example are formed in a manner which is consistent with conventional motion compensated prediction, while the high-pass sub-bands in the case of the 5-3 example are formed in a manner which is consistent with bi-directional motion compensated prediction. Unlike motion compensated prediction schemes, however, the present embodiment achieves perfect reconstruction without introducing the feedback loop which works against effective scalable compression in existing video compression standards.

As with conventional ID Wavelet transforms, the MCTWT may be recursively applied to the low pass sub-band to yield a hierarchy of temporal resolution levels. In preferred embodiments, this process is carried on three or four times so that the lowest frequency temporal sub-band has a frame rate which is one eighth or one sixteenth that of the original video sequence.

The temporal sub-bands produced by the MCTWT described above are further decomposed into spatial sub-bands by means of a spatial wavelet transform. Alternatively, other transforms such as the Discrete Cosine Transform (DCT) may be employed. These spatial transforms and their various forms and structure are well known to those skilled in the art. The result is the division of the original video sequence into a collection of spatio-temporal sub-bands.

The sequence of frames representing each of the spatio-temporal sub-bands is partitioned into code-blocks, where each code-block has a given spatial and temporal extent. In various embodiments of the invention, the extent of the code-blocks vary from sub-band to sub-band and may in one or more sub-bands be such as to span a single frame of the respective sub-band. In the preferred embodiment of the invention, code-blocks have relatively small spatial and temporal extents with a total volume of several thousand samples. For example, the spatial extent might be 32x32 or 16x16 with a temporal extent of 4 frames.

A completely independent embedded bit-stream is generated for every code-block. The embedded bit-stream for each code-block, B_i, may be truncated to any of a number of distinct lengths, R_i^s, such that each truncated bit-stream corresponds to a suitably efficient representation of the original sample values, in the rate-distortion sense. Efficient embedded block coding algorithms have been introduced for image compression in the form of the EBCOT algorithm, see D. Taubman, "EBCOT: Embedded block coding with optimised truncation," *ISO/IEC JTC 1/SC 29/WG1 N1020R*, October 1998 and the JPEG2000 image compression standard. A similar embedded block coding algorithm is used in the preferred embodiment of the invention. In various embodiments, however, any embedded entropy coding algorithm with similar properties may be used to generate the bit-streams, B_i.

One of the most important attributes of embedded block coding is that the separate embedded bit-streams may be optimally truncated so as to minimise distortion for a given constraint on the overall bit-rate. Moreover, the optimum truncation lengths for each code-block are non-decreasing functions of the target bit-rate. This means that the optimally truncated block bit-streams corresponding to an overall compressed bit-rate of R constitute a subset of the optimally truncated block bit-

streams for an overall compressed bit-rate of say 2R, and so on. This is exactly the scalability property required to enable the remote video browsing applications discussed above. Block truncation strategies which seek to minimise distortion for a given bit-rate, or minimise bit-rate for a given level of distortion are known as Post Compression Rate-Distortion (PCRD) optimisation strategies. PCRD optimisation schemes which optimise perceptual video quality are described in a separate application [visual99]. However, in various embodiments, different strategies may be employed to determine the block truncation points.

Compressed video bit-streams composed of independent embedded block bit-streams possess a number of additional attributes which are of great relevance to remote video browsing applications. Lower resolution versions of the video sequence may be reconstructed by discarding the code-blocks corresponding to higher spatial frequency sub-bands (for lower spatial resolution) or higher temporal frequency sub-bands (for lower temporal resolution). Each code-block has a well defined temporal extent (number of frames) so that improvements in reconstructed video quality may be selectively requested for different temporal segments by retrieving additional portions of the relevant embedded bit-streams. Finally, each code-block has a well defined spatial extent. Specifically, the set of code-blocks which are required to reconstruct a given spatial region of interest may be computed from the code-block dimensions, the support of the Wavelet basis functions, and the relevant motion parameters, thereby enabling interactive browsing modalities where the reconstructed video quality is selectively improved only in a particular spatial region of interest.

As already mentioned, PCRD optimisation schemes may be used to generate multiple sets of optimal block truncation points, corresponding to successively higher overall video bit-rates or quality levels. These are

grouped in quality layers within the final compressed video bit-stream: the first quality layer contains the truncated block bit-streams corresponding to the lowest bit-rate; subsequent quality layers contain the incremental contributions from each code block's embedded bit-stream which are required to achieve successively higher reconstructed video quality, at higher overall bit-rates. In this way, the final bit-stream consists of a collection of code-blocks which together span the entire spatial and temporal extent of the video sequence, whose embedded bit-streams are separated into quality layers. The final step is to augment the bit-stream with mechanisms to efficiently identify the layer contributions of each of the code blocks, so that all but the relevant portions of the bit-stream may be stripped to meet the needs of a particular user or application. Various embodiments may choose to represent this information in different ways.

Figure 2 illustrates an implementation of an embodiment of the present invention, as discussed above. It will be appreciated that the blocks shown in Figure 2 can be implemented by appropriate software and hardware in a computing device or system.

Referring to Figure 2, reference numeral 10 designates an arrangement which implements the MCTWT. This block processes a sequence of video frames, creating two or more temporal sub-bands. In the simplest case, there are two sub-bands, each having half the frame rate of the original video sequence. The low frequency sub-band contains good rendition of the video at half of the full frame rate, while the other (high frequency) sub-band contains details which are not predictable from the low frequency sub-band. As discussed above, the low frequency sub-band is preferably split into high and low frequency components. Such splitting then continues for a predetermined number of "temporal decomposition levels".

Block 11 processes each frame in each temporal sub-

band generated by Block 10, decomposing it into spatial sub-bands.

Block 12 groups the samples within each spatio-temporal sub-band produced by Block 11 into blocks spanning a given number of frames and having a given width and height within the sub-band. The blocks are coded independently from one another. Note that different embodiments and different sub-bands may use different block dimensions.

Block 13 generates an embedded bit-stream for each code-block. The embedded bit-stream has the property that as more of the bit-stream is received (made available to a decoder), the quality of the sub-band samples represented by the code-block improves progressively.

Block 14 determines an optimal number of bits from each code-block's embedded bit-stream to include in the final compressed data stream, subject to constraints on the compressed bit-rate or the quality of the reconstructed video. In a preferred embodiment, as discussed above, the PCRD optimisation algorithm block runs multiple times, creating a succession of layers within the compressed data stream, representing successively higher overall compressed bit-rate and video quality, such that each layer contains an optimal distribution of contributions from each code-block's embedded bit-stream. In the preferred embodiment, the PCRD algorithm should be sensitive to the relative importance of different spatio-temporal subbands to a human viewer, as a function of the activity in the video.

Block 15 assembles the code-block contributions, whose lengths are optimised by Block 14, into a final compressed data representation consisting of one or more video quality layers, together with all required information to recover the code-block contributions associated with each layer from the data stream. This identifying information allows subsets of the final compressed data stream to be readily extracted,

corresponding to successfully lower bit-rate and lower quality renditions of the video.

It will be appreciated that decompression is the inverse of compression, and a skilled person would be able to design an appropriate decompressor and implement an appropriate decompression method from the above description of the compression method and compressor. Invertibility is a direct consequence of the lifting structure used to implement the forward transformation operations associated with the compression process.

The provision of a highly scalable compressed video signal which can be transmitted over networks facilitates interactive remote browsing of compressed video. A further aspect of the present invention relates to a system and method for interactive remote browsing of scalable compressed video.

Figure 3 is a block diagram illustrating a server computing system 20 and a client computing system 21 connected by a network 22. As illustrated, the server 20 and the client 21 show the key elements in the system in accordance with the present embodiment for interactive remote browsing of scalable compressed video.

The client 21 includes a connection manager 24, a client cache 25, a decompressor 26 and a graphical user interface 27. The server 21 includes a connection manager 28, a sub-system 29 to maintain information concerning the status of the client's cache 25, and a file server 30.

The connection managers 24 and 28 manage the transfer of elements from the scalable compressed data stream between the server 21 and the client 22. The connection manager includes two components, one on the server 21 side and one on the client 22 side. It handles network transmission of the compressed data, acknowledgment of received data, monitoring of network conditions and so forth. Various embodiments may implement this block differently, depending upon the communication medium and the level of responsiveness required for effective user

interaction.

The client cache 25, stores compressed data transmitted by the server 21, implementing an advertised policy for discarding old data when the cache becomes too full. Various embodiments may use different caching strategies, subject only to the requirement that the server 21 should have some knowledge of the client's 22 cache capabilities from which to predict some or all of its state.

The client cache status tracking block 29 maintains a mirror image of the state of the client's cache 25, recording the identity of the compressed data elements which the client 22 is believed to have received and stored. This element replicates the advertised caching strategy employed by the client 22 to identify elements which may have been discarded from the client's cache. The client 22 need not necessarily discard these, but it must not discard elements which the server's cache tracking block 29 would predict as having been cached.

The file server 30 interacts with the scalable compressed video data stream and also with the client cache status tracking arrangement 29, using knowledge of the client's current region of interest within the video, to determine which elements of the compressed data stream should be transmitted to the client to maximise the perceived video quality. These elements are sent to the connection manager via block 29 which updates its model of the contents of the client's cache on the way.

The decompressor 26 interacts with the client cache 25 to recover elements of the compressed data stream which are relevant to the current time instant, resolution and spatial region over which the video is to be rendered into a user-defined view port, decompressing and rendering those elements as required.

The graphical user interface 27 interacts with the user, accepting and interpreting commands identifying the user's region of interest within the video (including the

spatial and temporal resolution of interest), sending this information periodically to the server and invoking the decompressor 26 to render the required video content, using all relevant information which is available in the cache. The graphical user interface 27 is also responsible for determining an appropriate delay between user requests and rendering of available data so as to give the server time to download at least some relevant elements from the compressed data stream. There is no requirement that all relevant compressed data be retrieved before rendering can commence. Compressed data which arrives late to the cache may be found useful if an interactive user instructs the system to go back over the same portion of the video sequence, rendering it again. In operation, the user specifies a spatio-temporal region of interest, optionally including information concerning the maximum spatial and temporal resolutions of interest. The region of interest may include the entire video sequence, or any subset thereof in space or time. The region of interest request is conveyed to the decompressor which sends requests to the client's cache and decompresses the returned code-block contributions as they arrive so as to satisfy the user's request. The cache, in turn, identifies the code-block contributions which must be requested from the server and forms an appropriate compact request to be sent to the server via the network connection manager. This server request is formed based upon the current contents of the cache and estimates of the state of the connection so as to enable streaming video delivery with a high degree of likelihood. Thus, if the region of interest is entirely new so that all information must be retrieved directly from the server, only the initial quality layers of the scalable bit-stream will be requested. On the other hand, if the region of interest has been visited before, a higher quality service will be sustainable in real time, since only those components which do not already exist in the cache need be

supplied by the server over the limited network connection.

The server optionally maintains a map of the contents of the client's cache. This enables the client to send its requests in a compact form, since the server is able to determine the actual code-block contributions which are required to satisfy the request without resending information already stored in the client's cache. The server organises the required code-block segments into an appropriate sequence, so as to facilitate real time decompression in the client as the material is streamed from the server.

In some embodiments of the invention, the server may generate requests automatically on behalf of the client during idle periods when the client is silent. For example, once a request has been satisfied, the server might send successively higher quality layers from code-blocks in the vicinity of the end point of the request, until further requests are received. In this way, the final frame in the user's requested region of interest will experience continually improving quality once the request has been satisfied. In other embodiments of the invention, alternative policies may be defined to determine the nature of the information which will be sent, if any, to a silent client.

In the embodiment described above with reference to Figure 2 it is possible to reconstruct the video at a reduced spatial resolution after discarding some of the higher spatial frequency sub-bands produced by a spatial wavelet transform. One significant drawback of this approach, however, is that the amount of motion information does not scale with the spatial resolution. For applications requiring spatial resolution scalability, an alternative embodiment of the invention is preferred, in which a multi-resolution transform is first used to decompose each original image frame, $x_t[m,n]$ into spatial sub-bands and the MCTWT is then applied to the sub-bands.

In the discussion which follows, the multi-resolution spatial transform is taken to be a two dimensional discrete wavelet transform (DWT). Figure 4 is a block diagram of a system for implementing spatial resolution scalable compression.

As shown in Figure 4, each original frame $x_k[m,n]$ is initially transformed into a reduced resolution image, $LL_k^1[m,n]$, together with three high frequency detail images, denoted $LH_k^1[m,n]$, $HL_k^1[m,n]$ and $HH_k^1[m,n]$. Each of these sub-band images has dimensions roughly half as large as those of the original image, and the total number of samples in all sub-bands is identical to the number of samples in the original image. We express this initial stage of spatial wavelet analysis as

$$\begin{pmatrix} LL_k^1 \\ LH_k^1 \\ HL_k^1 \\ HH_k^1 \end{pmatrix} = analysis(x_k) = \begin{pmatrix} analysis_{LL}(x_k) \\ analysis_{LH}(x_k) \\ analysis_{HL}(x_k) \\ analysis_{HH}(x_k) \end{pmatrix}$$

Iterative application of the analysis operator decomposes $LL_k^1[m,n]$ into an even lower resolution image, $LL_k^2[m,n]$ together with additional detail sub-bands, $LH_k^2[m,n]$, $HL_k^2[m,n]$ and $HH_k^2[m,n]$. Continuing in this way, a D level transform produces a total of $3D+1$ sub-bands, not counting the intermediate sub-bands, $LL_k^1[m,n]$ through $LL_k^{D-1}[m,n]$.

Importantly, each analysis stage is invertible. The inverse operators are known as synthesis stages and their operation is expressed as

$$LL_k^{d-1} = synthesis\left(LL_k^d, LH_k^d, HL_k^d, HH_k^d\right)$$

where $x_k[m,n]$ is conveniently identified with a level 0 sub-band, $LL_k^0[m,n]$.

In the preferred embodiment of the invention, the MCTWT techniques developed previously are first applied directly to the lowest resolution sub-band frames, LL_k^D as shown in Figure 4. This MCTWT uses its own motion information, tailored to match the spatial resolution of the LL_k^D frames. In the specific example of the motion compensated temporal 5-3 wavelet transform, equations (3) and (4) become equations (5) and (6) below. Note that the prefixes "H-" and "L-" are used here to identify high- and low-pass temporal sub-bands, respectively. The superscript, D, in the motion compensated mapping operators, $W_{i,j}^D$ associate them with the lowest spatial resolution frames in the D-level DWT.

(5) $$H-LL_k^D[m,n]=LL_{2k+1}^D[m,n] -\frac{1}{2}\left\{\left(W_{2k,2k+1}^D(LL_{2k}^D)\right)[m,n]+\left(W_{2k+2,2k+1}^D(LL_{2k+2}^D)\right)[m,n]\right\}$$

and

(6) $$L-LL_k^D[m,n]=LL_{2k}^D[m,n] +\frac{1}{4}\left\{\left(W_{2k-1,2k}^D(H-LL_{k-1}^D)\right)[m,n]+\left(W_{2k+1,2k}^D(H-LL_k^D)\right)[m,n]\right\}$$

Each successive set of detail sub-bands, LH_k^d, HL_k^d and HH_k^d collectively form a "resolution level" R_k^{d-1} where d runs from D down to 1. Resolution level R_k^{d-1} holds the

information required to augment the spatial resolution from LL_t^d to LL_t^{d-1} (through spatial sub-band synthesis). As indicated in Figure 4, each resolution level has its own MCTWT, with its own motion information, tailored to match the spatial resolution of the LL_t^{d-1} frames. The associated motion compensated mapping operators are denoted $W_{i,j}^{d-1}$.

The MCTWT for any given resolution level, R_t^{d-1} is similar in principle to that described hitherto, with the exception that effective motion compensation should be performed in the domain of the spatially synthesised frames, LL_t^{d-1}. The temporal lifting steps are applied directly to the detail sub-bands, but whenever motion compensation is required, the detail sub-bands in the resolution level are jointly subjected to spatial synthesis to recover a baseband image with the same resolution as LL_t^{d-1}; motion compensation is applied in this domain, followed by sub-band analysis, which yields the motion compensated detail sub-bands. To minimise spatial aliasing effects which can interfere with the quality of the motion compensation and hence the compression efficiency of the overall transform, some information is borrowed from the lower resolution frames, LL_t^d as shown in Figure 4. This information is available at both the compressor and the decompressor, since lower spatial resolutions can be reconstructed first, without reference to higher spatial resolutions.

For illustrative purposes, consider again the motion compensated 5-3 wavelet transform. The lifting steps for resolution level R_t^{d-1} employ motion compensation operators, $W_{i,j}^{d-1}$ applying them in accordance with equations (7) and (8) below.

$$\begin{pmatrix} H-LH_k^d \\ H-HL_k^d \\ H-HH_k^d \end{pmatrix} = \begin{pmatrix} LH_{2k+1}^d \\ HL_{2k+1}^d \\ HH_{2k+1}^d \end{pmatrix}$$

$$(7) \quad -\frac{1}{2}\begin{pmatrix} analysis_{LH}\left(W_{2k,2k+1}^{d-1}\left(synthesis(LL_{2k}^d, LH_{2k}^d, HL_{2k}^d, HH_k^d)\right)\right) \\ analysis_{HL}\left(W_{2k,2k+1}^{d-1}\left(synthesis(LL_{2k}^d, LH_{2k}^d, HL_{2k}^d, HH_k^d)\right)\right) \\ analysis_{HH}\left(W_{2k,2k+1}^{d-1}\left(synthesis(LL_{2k}^d, LH_{2k}^d, HL_{2k}^d, HH_k^d)\right)\right) \end{pmatrix}$$
$$-\frac{1}{2}\begin{pmatrix} analysis_{LH}\left(W_{2k+2,2k+1}^{d-1}\left(synthesis(LL_{2k+2}^d, LH_{2k+2}^d, HL_{2k+2}^d, HH_{2k+2}^d)\right)\right) \\ analysis_{HL}\left(W_{2k+2,2k+1}^{d-1}\left(synthesis(LL_{2k+2}^d, LH_{2k+2}^d, HL_{2k+2}^d, HH_{2k+2}^d)\right)\right) \\ analysis_{HH}\left(W_{2k+2,2k+1}^{d-1}\left(synthesis(LL_{2k+2}^d, LH_{2k+2}^d, HL_{2k+2}^d, HH_{2k+2}^d)\right)\right) \end{pmatrix}$$

and

$$\begin{pmatrix} L-LH_k^d \\ L-HL_k^d \\ L-HH_k^d \end{pmatrix} = \begin{pmatrix} LH_{2k}^d \\ HL_{2k}^d \\ HH_{2k}^d \end{pmatrix}$$

$$(8) \quad +\frac{1}{4}\begin{pmatrix} analysis_{LH}\left(W_{2k-1,2k}^{d-1}\left(synthesis(\overline{H-LL_{k-1}^d}, H-LH_{k-1}^d, H-HL_{k-1}^d, H-HH_{k-1}^d)\right)\right) \\ analysis_{HL}\left(W_{2k-1,2k}^{d-1}\left(synthesis(\overline{H-LL_{k-1}^d}, H-LH_{k-1}^d, H-HL_{k-1}^d, H-HH_{k-1}^d)\right)\right) \\ analysis_{HH}\left(W_{2k-1,2k}^{d-1}\left(synthesis(\overline{H-LL_{k-1}^d}, H-LH_{k-1}^d, H-HL_{k-1}^d, H-HH_{k-1}^d)\right)\right) \end{pmatrix}$$
$$+\frac{1}{4}\begin{pmatrix} analysis_{LH}\left(W_{2k-1,2k}^{d-1}\left(synthesis(\overline{H-LL_k^d}, H-LH_k^d, H-HL_k^d, H-HH_k^d)\right)\right) \\ analysis_{HL}\left(W_{2k-1,2k}^{d-1}\left(synthesis(\overline{H-LL_k^d}, H-LH_k^d, H-HL_k^d, H-HH_k^d)\right)\right) \\ analysis_{HH}\left(W_{2k-1,2k}^{d-1}\left(synthesis(\overline{H-LL_k^d}, H-LH_k^d, H-HL_k^d, H-HH_k^d)\right)\right) \end{pmatrix}$$

The terms, $\overline{H-LL_k^d}$ appearing in equation (8) above, are designed to make the synthesised frames, to which the motion compensation is actually applied, resemble those which would appear if the relevant lifting steps were applied directly at video resolution, LL_k^{d-1}. For the entire transform to remain invertible, it is essential that $\overline{H-LL_k^d}$ be based only on information from video resolution, LL_k^d.

In some embodiments, these terms can all be set to 0, at the expense of some uncompensated aliasing effects as the detail sub-bands are synthesised, motion compensated and

re-analysed. In other embodiments, $\overline{H-LL_k^d}$ may be obtained by spatially synthesising sub-bands $H-LL_k^0$, $H-LH_k^0$, ..., $H-HH_k^{d-1}$.

In the preferred embodiment, $H-LL_k^d$ is obtained by interpolating the LL_k^d frames (these are available to both the compressor and the decompressor, after reconstruction) through one stage of spatial synthesis (with all detail sub-bands set to 0), applying the motion compensated temporal lifting steps in this interpolated domain, and then recovering $H-LL_k^d$ through spatial DWT analysis of the temporal sub-band, $H-LL_k^{d-1}$ discarding the detail sub-bands produced by the analysis operator. The advantage of this particular embodiment over the previous one is that the motion compensated lifting steps used to derive $H-LL_k^{d-1}$ and thence $H-LL_k^d$ are able to exploit all of the motion information associated with the higher density mapping operator, $W_{i,j}^{d-1}$.

The ideas expressed above extend naturally to arbitrary temporal wavelet transforms with any number of lifting steps and any number of levels of temporal decomposition; such generalisations will be apparent to those skilled in the art. The nature of the final three dimensional transform is always such that the video frames may be recovered at any given video resolution, LL_k^d using only the motion information corresponding to the operators, $\{W_{i,j}^0\}$ through $\{W_{i,j}^d\}$. This allows the total amount of motion information (e.g. the total number of block motion vectors, mesh node velocities, etc.) to scale with the resolution of interest within a final compressed data stream.

Although the above discussion is restricted to the context of a monochrome video source, this is only for simplicity of explanation. Generalisation to multiple components, including those required for colour video, should be apparent to those skilled in the art.

Although the discussion above is concerned with the

transformation and compression of motion picture information, it should be apparent that the invention is applicable to other applications in which sequences of images are to be compressed. Examples of particular interest include volumetric medical imagery and the images used by image-based three dimensional rendering applications.

It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present invention as shown in the specific embodiment without departing from the spirit or scope of the invention as broadly described. The present embodiment is, therefore, to be considered in all respects to be illustrative and not restrictive.

CLAIMS

1. A method of compressing a video sequence to produce a compressed video signal, including the steps of forming the video sequence into an input signal, decomposing the input signal into a set of temporal frequency bands, in a manner which exploits motion redundancy, and applying scalable coding methods to the decomposed input signal to generate a scalable compressed video signal.

2. A method in accordance with claim 1, wherein the step of decomposing the input signal comprises the further step of further decomposing the temporal frequency bands into spatial frequency bands, to result in spatio-temporal frequency bands.

3. A method in accordance with claim 1 or claim 2, wherein the step of decomposing the input signal into a set of temporal frequency bands includes the steps of separating the input signal into even and odd indexed frames sub-sequences, and applying a sequence of motion compensated lifting operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa in a manner which is sensitive to motion.

4. A method in accordance with claim 3, including the further step of identifying the final updated even frame sub-sequence with the low temporal frequency band and the final updated odd frame sub-sequence with the high temporal frequency band.

5. A method in accordance with claim 3 or claim 4, including the further step of further decomposing the low temporal frequency band for a predetermined number of decomposition levels, by recursively applying the steps of claim 3.

6. A method in accordance with any one of claims 3 to 5, wherein the lifting operations are identical to those required to affect any of a range of efficient 1-D wavelet transforms, with the addition that a motion compensated

version of the even sub-sequence is used in performing the lifting operation which updates the odd sub-sequence.

7. A method in accordance with claim 6, wherein a motion compensated version of the odd sub-sequence is used in performing the lifting operation which updates the even sub-sequence.

8. A method in accordance with claim 6 or claim 7, wherein block based motion models are used.

9. A method in accordance with claim 6 or claim 7, wherein mesh based motion models are used.

10. A method in accordance with any one of the preceding claims, wherein the step of applying coding methods includes the step of partitioning the frequency bands into code-blocks, with a given size in each of one or more predetermined dimensions.

11. A method in accordance with claim 10, wherein the frequency bands are the spatio-temporal frequency bands of claim 2, and the partitioning step includes partitioning the spatio-temporal frequency bands into 3-D code-blocks with a pre-determined size in each of the spatial dimension and the temporal dimension.

12. A method in accordance with claim 11, comprising the further step of forming an independent embedded block bit-stream for each of the code-blocks.

13. A method in accordance with claim 12, comprising the further step of separately truncating the embedded block bit-streams in order to maximise the quality of the represented video sequence subject to a given constraint on bit-rate.

14. A method in accordance with claim 12, comprising the further step of separately truncating the embedded block bit-streams to minimise bit-rate subject to a constraint on the quality of the represented video sequence.

15. A method in accordance with claim 13 or 14, wherein the embedded block bit-streams are organised in an interleaved fashion to provide quality layers, each successive quality layer consisting of the incremental

contributions for each code-block bit-stream which are required to augment the information contained in previously quality layers so as to embody the truncated block bit-streams corresponding to the next higher quality or bit-rate constraint.

16. A method in accordance with any one of the preceding claims, wherein the step of forming the video sequence into an input signal comprises the step of providing the video sequence as the input signal.

17. A method in accordance with claim 1, wherein the step of forming the video sequence into an input signal, includes the step of first decomposing each original image frame into spatial sub-bands, and providing the sequence of decomposed image frames as the input signal.

18. A method in accordance with claim 17, wherein the spatial subbands from different resolution levels have different motion fields, each motion field having an information density commensurate with the corresponding spatial resolution.

19. A method in accordance with claim 17 or claim 18, wherein the step of decomposing the input signal into a set of temporal frequency bands includes the steps of separating the input signal into even and odd indexed frame sub-sequences, and applying a sequence of motion compensated lifting operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa in a manner which is sensitive to motion, the motion compensation within any or all of the lifting steps being performed jointly for all detail subbands of a single resolution level.

20. A method in accordance with claim 19, in which the joint motion compensation of detail subbands belonging to a resolution level consists of a spatial wavelet synthesis step, followed by a motion compensation step, applied to the synthesised frames, and then a spatial subband analysis step to recover the motion compensated detail subbands.

21. A method in accordance with claim 20, in which the spatial wavelet synthesis step which precedes each motion compensation step within a resolution level is augmented using information derived from the lower spatial resolution frames which can be reconstructed by a decompressor without reference to the information from the resolution level in question.

22. A method of providing for interactive remote browsing of compressed digital video signals, comprising the steps of:

making available to a network, by way of a server computing system connected to the network, a scalable compressed signal representing a video sequence, selecting, by way of a client computing system connected to the network, a portion of the scalable compressed signal for decompression and viewing as a video sequence, monitoring the choice of selection by the client, and selecting, by way of the client, a further portion of scalable compressed signal of interest to the client.

23. A method in accordance with claim 22including the further step of the server monitoring the selection of the portion of the signal by the client, whereby when the client requests a further portion of signal in a particular region of interest, the server is able to provide the most appropriate portion of signal to the client.

24. A method in accordance with claims22 or 23 wherein the signal is scalable in a plurality of dimensions, and the method includes the further step of the client selecting the portion relating to the dimension of the region of interest.

25. A method in accordance with claim 24 wherein the dimensions include spatial areas of the video and temporal areas.

26. A method in accordance with any one of claims 22 to

25 comprising the further step of the client computing system caching received portions of the scalable compressed signal, whereby the further portions of the scalable compressed signal may be added to the cached portions for decompression to produce video in the region of interest.

27. A method in accordance with claim 26, wherein the server computing system keeps track of the data which has been cached by the client computing system and uses this information to determine the most appropriate information to transmit to the client computing system, so as to more effectively utilise the network resources.

28. A method in accordance with any one of claims 22 to 27 wherein the scalable compressed signal is compressed by the method of any one of claims 1-21.

29. A method of compressing a video sequence to produce a compressed video signal, including the steps of decomposing the video sequence into a set of temporal frequency side bands by separating the video sequence into even and odd indexed frame sub-sequences, and applying a sequence of motion compensated lifting operations to alternately update the odd frame-sequence based on the even frame-sequence and vice versa in a manner which is sensitive to motion, and applying coding methods to generate a compressed video signal.

30. A method in accordance with claim 29, further comprising the preceding step of decomposing each original image frame into spatial subbands and then applying the steps of claim 29 to the video sequence decomposed into spatial subbands.

31. A system for compressing a video sequence to produce a compressed video signal, the system comprising means for forming the video sequence into an input signal, means for decomposing the input signal into a set of temporal frequency bands, in a manner which exploits motion redundancy, and a means for applying a scalable coding method to the decomposed input signal to generate a

scalable compressed video signal.

32. A system in accordance with claim 31, further including means for decomposing the temporal frequency bands into spatial frequency bands, to result in spatio-temporal frequency bands.

33. A system in accordance with claim 31 or claim 32 wherein the means for decomposing the input signal into a set of temporal frequency bands includes a lifting means, arranged to separate the input signal into even and odd indexed frame sub-sequences, and apply a sequence of motion compensated lifting operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa in a manner which is sensitive to motion.

34. A system in accordance with claim 33, wherein the lifting means includes means for identifying the final updated even frame sub-sequence with the low temporal frequency band and the final updated odd frame sub-sequence with the high temporal frequency band.

35. A system in accordance with claim 33 or claim 34, wherein the lifting means is arranged to further decompose the low temporal frequency band for a predetermined number of decomposition levels in a recursive manner.

36. A system in accordance with any one of claims 33 to 35, wherein the lifting operations are identical to those required to affect any of a range of efficient 1-D wavelet transforms, with the addition that a motion compensated version of the even sub-sequence is used in performing the lifting operation which updates the odd sub-sequence.

37. A system in accordance with claim 36, wherein a motion compensated version of the odd sub-sequence is used in performing the lifting operation which updates the even sub-sequence.

38. A system in accordance with claims 36 or 37, wherein block based motion models are used.

39. A system in accordance with claim 36 or 37, wherein mesh based motion models are used.

40. A system in accordance with any one of claims 31 to 39, wherein the scalable coding method means includes means for partitioning the frequency bands into code-blocks, with a given size in each of one or more predetermined dimensions.

41. A system in accordance with claim 40 wherein the frequency bands are the spatio-temporal frequency bands referred to in claim 32 and the means for partitioning is arranged to partition the spatio-temporal frequency bands into 3-D code-blocks with a pre-determined size in each of the spatial dimension and the temporal dimension.

42. A system in accordance with claim 41 the partitioning means further being arranged to form an independent embedded block bit-stream for each of the code-blocks.

43. A system in accordance with claim 42, the partitioning means further being arranged to truncate the embedded block bit-streams in order to maximise the quality of the represented video sequence subject to a given constraint on bit-rate.

44. A system in accordance with claim 42, the partitioning means further being arranged to separately truncate the embedded block bit-streams to minimise bit-rate subject to a constraint on the quality of the represented video sequence.

45. A system in accordance with claim 43 or claim 44, wherein the embedded block bit-streams are organised in an interleaved fashion to provide quality layers, each successive quality layer consisting of the incremental contributions for each code-block bit-stream which are required to augment the information contained in previously quality layers so as to embody the truncated block bit-streams corresponding to the next higher quality or bit-rate constraint.

46. A system in accordance with any one of claims 31 to 45, the means for forming the video sequence into an input signal being arranged to provide the video sequence as the input signal.

47. A system in accordance with claim 31, wherein the means for forming the video sequence into an input signal is arranged to first decompose each original image frame into spatial sub-bands, and provide the sequence of decomposed image frames as the input signal.

48. A system in accordance with claim 47, wherein the spatial subbands from different resolution levels have different motion fields, each motion field having an information density commensurate with the corresponding spatial resolution.

49. A system in accordance with claim 47 or claim 48, wherein the means for decomposing the input signal into a set of temporal frequency bands comprises a lifting means arranged to separate the input signal into even and odd indexed frame sub-sequences, and apply a sequence of motion compensated lifting operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa in a manner which is sensitive to motion, the motion compensation within any or all of the lifting steps being performed jointly for all detail subbands of a single resolution level.

50. A system in accordance with claim 49 wherein the joint motion compensation of detail subbands belonging to a resolution level consists of a spatial wavelet synthesis step, followed by a motion compensation step, applied to the synthesised frames, and then a spatial subband analysis step to recover the motion compensated detail subbands.

51. A system in accordance with claim 50, wherein the spatial synthesis step which precedes each motion compensation step within a resolution level is augmented using information derived from the lower spatial resolution frames which can be reconstructed by a decompressor without reference to the information from the resolution level in question.

52. A system for providing for interactive remote browsing of compressed digital video signals, comprising:

a server computing system arranged to be connected to a network and including means for providing a scalable compressed signal representing a video sequence for distribution on the network to a client computing system,

a client computing system, including selection means for selecting a portion of the scalable compressed signal, a decompressor for decompressing the scalable signal for viewing as a video sequence, the selection means being arranged to select further portions of the scalable signal relating to regions of interest of the client.

53. A system in accordance with claim 52 wherein the client includes a cache arranged to store portions of the scalable signal and the server includes a cache monitor arranged to monitor the client to enable the server to determine what portions of the signal are stored in the client cache, whereby the server, on request of the client for a region of interest video, may send the client only the further portions of signal required to reproduce the region of interest.

54. An apparatus for compressing a video sequence to produce a compressed video signal, comprising means for decomposing the video sequence into a set of temporal bands, and separating the input signal into even and odd index frame sub-sequences and applying a sequence of motion compensated lifting operations to alternately update the odd frame sub-sequence based upon the even frame sub-sequence and vice versa in a manner which is sensitive to motion.

55. A method of decompressing a video signal compressed in accordance with the method steps of any one of claims 1 to 21, comprising the steps of applying method steps which are the inverse of the steps of any one of claims 1 to 21 or 29 and 30.

56. A decompressor system including means for applying the method of claim 55.

57. A computer program arranged, when loaded onto a computing system, to control the computing system to

implement the method of any one of claims 1 to 21 or 29 and 30.

58. A computer readable medium, arranged to provide a computer program in accordance with claim 57.

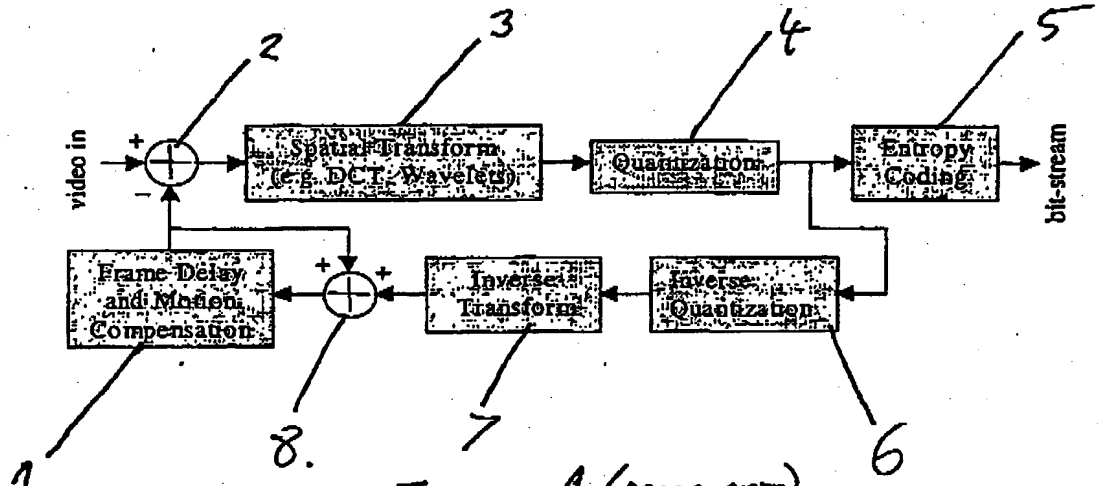


FIGURE 1 (PRIOR ART)

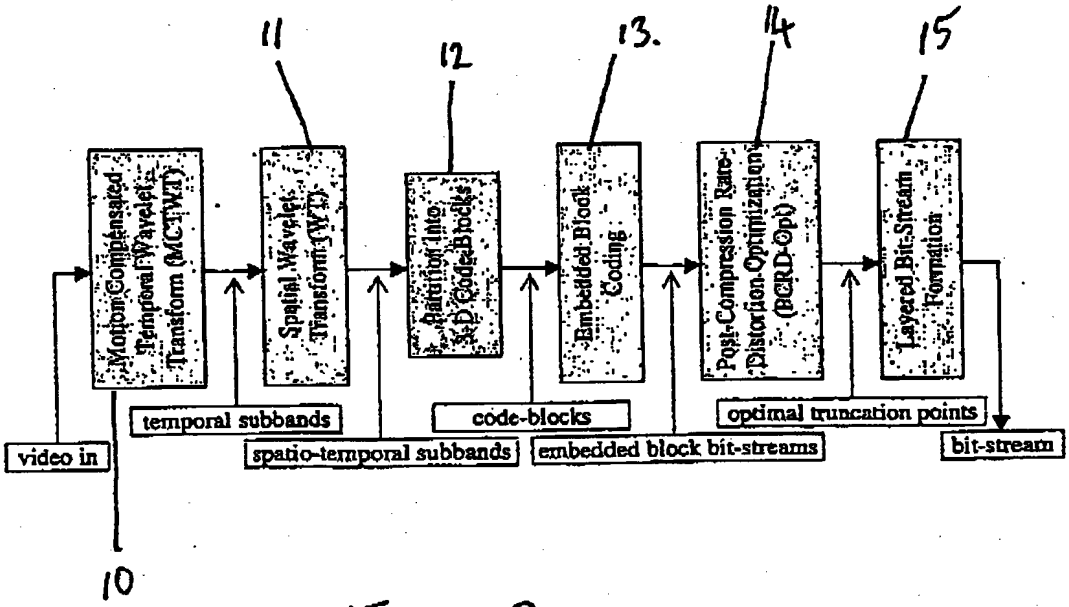


FIGURE 2

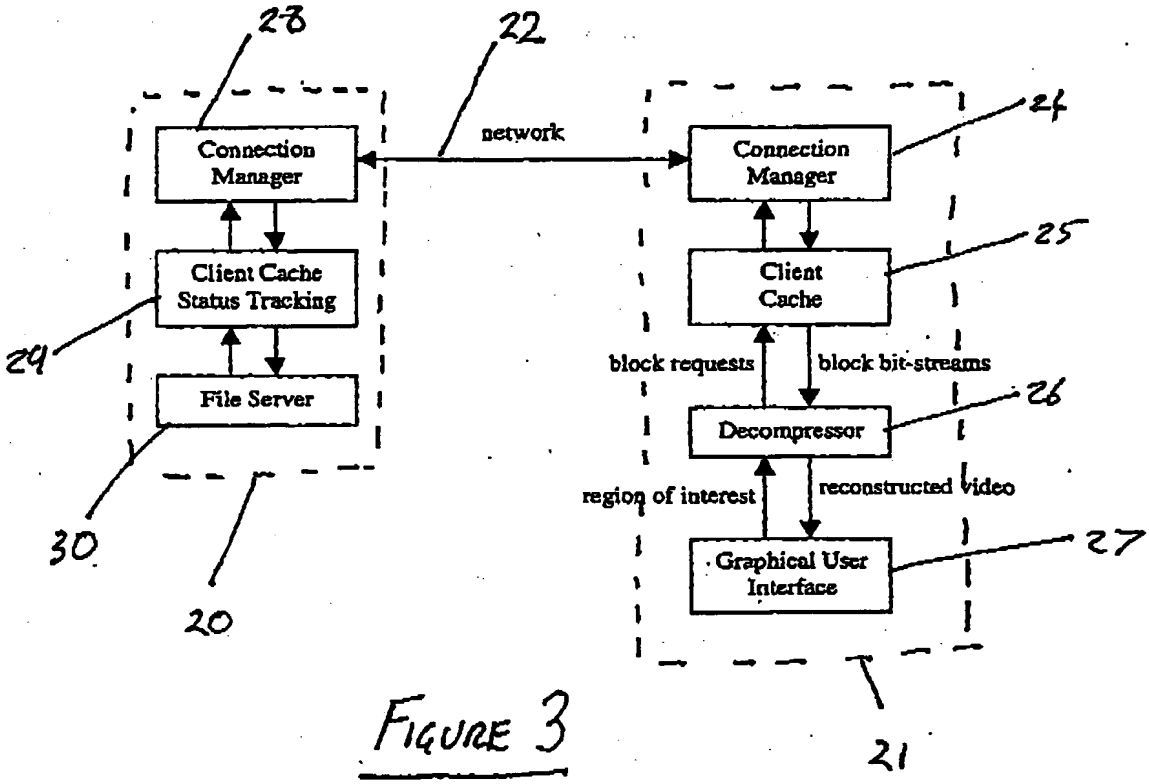


FIGURE 3

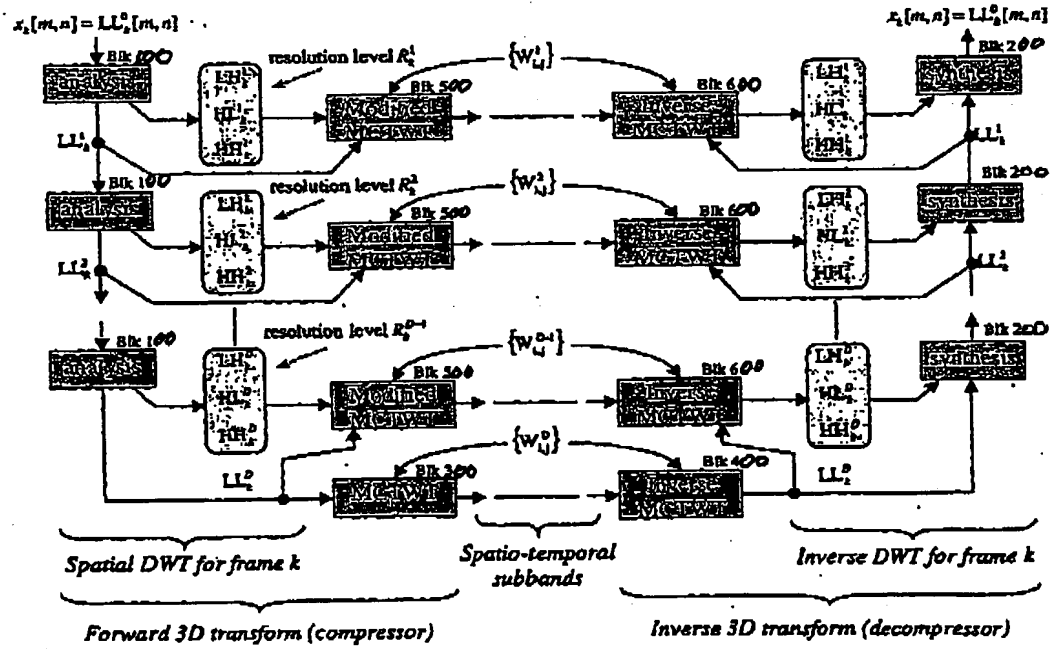


FIGURE 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/01660

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: G06T 9/00, H04N 7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT,CiteSeer (image, compression, scalable, temporal frequency bands)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00/74385 (Gu et al.) 7 December 2000 Abstract, pages 2-4, page 22 lines 6 - 22	1,2,31,32
X	"Scalable Object-Based 3-D Subband/Wavelet Coding of Video" (Han et al.) 30 October 1998 IEEE Trans. on Circuits and Systems for Video Technology 30 pages	1,2,31,32,16,31,32,46
X	"Very Low Bit-Rate Embedded Video Coding with 3D Set Partitioning in Hierarchical Trees (3D SPIHT)	1,2,31,32
Y	Sections 2.2,4.1,5	22,52

[X] Further documents are listed in the continuation of Box C [X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
26 February 2002

Date of mailing of the international search report 06 MAR 2002

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaustralia.gov.au
Facsimile No. (02) 6285 3929

Authorized officer

DALE E. SIVER
Telephone No : (02) 6283 2196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/01660

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5659363 (Wilkinson) 19 August 1997 column 3 lines 24-28	1,31
Y	"Spatio-Temporal Segmentation and General Motion Characterization for Video Indexing and Retrieval" (Fablet et al.) June 1999 10th DELOS Workshop on Audio-Visual Digital Libraries Paragraph bridging pages 4 and 5	22,52

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU01/01660

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member			
WO	200074385	AU	200052942		
US	5659363	GB	2286740	JP	7284108

END OF ANNEX